Problem Studied

The problem studied was to significantly enhance the capability and usability of online software and data repositories. To this end, a variety of approaches were employed:

- Development of mechanisms for location of resource replicas, with a guarantee of consistency and integrity for those replicas;
- Development of new interfaces for user interaction with software and data repositories, enhancing the user's ability to identify appropriate software and/or data for his needs;
- Development of active repository technologies, for example:
    - repositories that build specialized software components on demand,
    - repositories that deliver portable, mobile software components that can be integrated with applications without rebuilding them,
    - applications programming interfaces to network-based computational servers, which deliver results to common mathematical computations on demand.

Summary of the Most Important Results

The Resource Cataloging and Distribution System (RCDS) was designed to facilitate scalable distribution of resources to multiple sites for efficiency and redundancy, while assuring authenticity, integrity, and consistency of the resources and metadata. Software and data repositories can use RCDS as a substrate to maintain replicas of their resources, while supporting a decentralized management paradigm. Users of those repositories benefit from more reliable and efficient access. RCDS employs digital signatures to protect against malicious or accidental modification of codes and data provided through its service.

The SONAR proximity estimation system aids a client in selecting a nearby replica of a resource, by providing a relative ordering of replica locations according to their network proximity to the client. This improves response time and provides better utilization of the network.

Another package was developed to provide a suite of software access control and licensing tools. This package allows software maintainers to select from a variety of user authentication methods to provide access control to their software, thus protecting their intellectual property. In addition, the package allows repository maintainers to require registration and/or a license agreement before the software can be downloaded.

In the area of improved interfaces for software repositories, the Approximation Wizard is an example of "data-driven searching;" the user supplies a data file and the Approximation Wizard deduces its format and runs a battery of tests over it, looking for distinctive features such as monotonicity, noise, strong frequency components, and so on. After confirming those diagnostics with the user, it selects specific algorithms that are appropriate for the data, supplying graphical displays of the results.

REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	05/30/00	Final, 9/28/95 - 9/30/99

4. TITLE AND SUBTITLE
Mechanisms for Adaptable and Efficient Information Retrieval Clients and Servers

5. FUNDING NUMBERS
DAAH04-95-1-0595

6. AUTHOR(S)
Keith Moore, author

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Tennessee
104 Ayres Hall
Knoxville, TN 37996-1301

8. PERFORMING ORGANIZATION REPORT NUMBER F-1

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 34972.4-EL

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This research project was conducted with the aspiration of significantly enhancing the capability and usability of online software and data repositories. The Resource Cataloging and Distribution System (RCDS) was designed to facilitate scalable distribution of resources to multiple sites for efficiency and redundancy, while assuring authenticity, integrity and consistency of the resources and metadata. The RCDS software is also being used to support a variety of ongoing research projects, including projects related to software reuse as well as other applications of highly scalable resolution and replication services. Some of these research projects are: Program Builder, Netbuild, Scalable Networked Information Processing Environment (SNIPE), MPI-Connect, HARNeSS and NetSolve.

Experience derived from the use of the Resource Catalog resolution subsystem of RCDS is being used to design a highly scalable resolution system for Uniform Resource Names (URNs) and other kinds of URIs. URNs will serve as very long-term stable names for Internet-accessible resources, in contrast to URLs, which change if the resource moves from one host to another, or if the domain or repository is reorganized.

14. SUBJECT TERMS
software and data repositories
user interaction interfaces
assurance of authenticity, integrity, consistency

15. NUMBER OF PAGES
8

16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

The Guide to Available Mathematical Software (GAMS) is a cross-index and virtual repository of public-domain and commercial software components. GAMS provides enhanced indexing and content-specific search capabilities for a superset of the netlib repository.

The Matrix Market is a database of artifacts used in testing and evaluating algorithms and software for numerical linear algebra. The system provides search facilities for matrix data, visualizations of matrix structure, and facilities for on-demand generation of matrices to meet user-specified criteria.

The Digital Library of Mathematical Functions provides online reference data for the special functions of applied mathematics. The emphasis is on development of active components, such as interactive visualizations, identities that can be cut-and-pasted into computer algebra systems, and tables of certified data that are generated on demand. This is an attempt to address the challenge of providing effective search techniques for specialized mathematic reference data, without extensive auxiliary text.

Several of these repositories have adopted the Basic Interoperability Data Model (BIDM) developed by the Reuse Library Interoperability Group (RIG), of which the project participants have been active members. The BIDM is an IEEE standard that specifies a minimal set of catalog information that a software repository should provide about its software resources in order to interoperate with other repositories. The BIDM is expressed in terms of an extended entity-relationship data model that defines classes for assets (the software entities), the individual elements making up assets (i.e., files), reuse libraries (i.e., repositories) that provide assets, and organizations that develop and manage libraries and assets. The BIDM helps provide semantic interoperability between different repositories.

A toolkit known as Repository in a Box (RIB) enables software repository developers to create and maintain software catalog records using the BIDM, to exchange those records as well as software resources themselves with other repositories, and to provide a user interface for their software catalog. RIB software catalogs are maintained and accessed using a WWW interface. RIB administrative functions available through this interface allow maintenance of software catalog records, exchange of catalog records with other repositories, administration of browsing and searching interfaces, discipline-oriented and site-specific customization, file uploading, and replication. Multiple repositories may be maintained at the same site using a single RIB installation.

Java has proven to be very useful in these efforts. For GAMS, we have been able to develop a complete portable user interface in Java, which provides a variety of capabilities not present in the hypertext version. However, for the Matrix Market, we designed a series of applets that allow for the generation of matrix test data in a user's browser.

This work has demonstrated the great promise of Java in the evolution of software repositories. Among its key features is the possibility of truly mobile mathematical software applications and components. The use of such components could have a significant impact on how such software is distributed and used. To explore this further, we undertook a systematic study of the suitability of Java for numerical computation. This study paints a mixed picture of the capabilities of Java and extant Java Virtual Machines (JVMs). Among the problems are serious performance barriers, a lack of key facilities enabling efficient scientific software development, and the lack

of a credible base of software components that can support serious numerical computation. To help make Sun and others aware of needs of the high performance computing community we, in collaboration with Sun, Intel, IBM, The MathWorks, NAG, Visual Numerics, Syracuse University, the University of California at Berkeley, the University of North Carolina, Indiana University, and others, formed the Java Grande Forum. Ron Boisvert and Roldan Pozo of the National Institute of Standards Technology (NIST) chair the Forum's Numerics Working Group. This group is working to improve the Java language and its environment for numeric-intensive computations through both changes in the Java language and its semantics, as well as the development of standard class libraries for core numerical functions.

Another key technology for search and retrieval, as well as for interchange of active mathematical data, is the development of ontologies and related semantic-based markup systems for mathematical data. An example of this work's importance is its use in the Digital Library of Mathematical Functions. The W3C's adoption of MathML has provided an important first step in this direction; however, the semantic markup in MathML is too weak to support the description of higher mathematical concepts necessary for the interchange of data among computer algebra systems. We are working with the European OpenMath Consortium and the North American OpenMath Initiative (NAOMI) to develop extensions to MathML, which provide rich semantic-based markup for mathematical data, as well as associated tools.

The RCDS software is being used to support a variety of ongoing research projects, including projects related to software reuse, as well as other applications of highly scalable resolution and replication services:

- The Program Builder is a tool that assembles computer programs out of source code components obtained from repositories. It can download a description of the program describing the components needed to build a program and instructions for building it, retrieve the necessary components from various locations, verify the authenticity and integrity of each, configure them according to the target environment, compile them, and optionally install them. The program builder uses RCDS to store and retrieve catalog information for the various packages, and uses signed RCDS catalog information to verify the authenticity and integrity of the packages it uses.

- Similar to the Program Builder, Netbuild is a link-editor that allows seamless access to remote repositories of pre-compiled subroutine libraries, such as the mathematical libraries in the Netlib repository. Netbuild uses RCDS to catalog its subroutine libraries, to identify which versions of the libraries are appropriate for the target platform, and to verify the authenticity and integrity of the libraries before they are linked.

- The Scalable Networked Information Processing Environment (SNIPE) is a large-scale, fault-tolerant, distributed computing environment layered on top of RCDS. It uses the Resource Catalog component of RCDS to store metadata about computing resources, processes, resource managers, multicast groups, programs, data files, and checkpointed programs. The SNIPE environment utilizes RCDS's fault tolerance to allow it to survive multiple host failures. Processes can migrate from one host to another for load balancing or to avoid host failures; their communications peers will automatically re-establish communications at the new location. As in other projects, signed RCDS metadata is used to verify authenticity and integrity of programs, data files, and checkpoint files. RCDS

metadata is also used to list communications ports by which a process can be reached, thus allowing a peer to choose the best available path when establishing a connection.

- The Resource Catalog portion of RCDS is also used as a substrate in implementations of MPI_Connect. MPI_Connect is being used by CEWES and ASC DoD MSRCs. It has also been used by the High Performance Computing Center in Stuttgart, Germany, and the Paderborn Center for Parallel Computing (P2C).

- Harness is another distributed computing environment that utilizes RCDS in a similar fashion to SNIPE. Harness uses the Resource Catalog component of RCDS to provide a very flexible and highly customizable metacomputing environment through the extensive use of RCDS-cataloged plug-in modules. These allow a "building block" approach to the construction of one or more "virtual machines," consisting of computing, communications resources and software repository resources, to be employed in a computation. Harness also uses RCDS to facilitate merging, splitting of virtual machines and communications between virtual machines.

- NetSolve provides a foundation for network-based computational servers, allowing users to access computational resources such as hardware and software, to be distributed across the network. NetSolve is investigating the use of RCDS as a catalog for computational resources.

Experience derived from the use of the Resource Catalog resolution subsystem of RCDS, as well as from similar systems, is being used to design a highly scalable resolution system for Uniform Resource Names (URNs) and other kinds of URIs. URNs will serve as very long-term stable names for Internet-accessible resources, in contrast to URLs, which change if the resource moves from one host to another, or if the domain or repository is reorganized. URNs have been standardized by the Internet Engineering Task Force (IETF).

The IETF is also considering standardization of a general purpose web replication/caching infrastructure and general-purpose name resolution systems (tentatively known as RESCAP), based on experience with the RCDS work. Anticipated applications of this technology include the use of RCDS-like resolution services for content or protocol negotiation between Internet clients and servers. An example of such an application is determining the capabilities of an electronic mail recipient's user agent (e.g. "Does joe@xyz.com accept powerpoint?"), or mapping an E.164 telephone number into one or more recipient-specified service locations, which provide routing for a voice, fax, or pager call to that number.

The RIB toolkit is currently deployed and in use at three of the Department of Defense Major Shared Resource Centers (MSRCs) for setting up repositories for several of the DoD Computational Technology Areas (CTAs) as part of the High Performance Computing Modernization Program. Other sites that are using RIB to set up software repositories include the NASA Earth and Space Sciences program and the NSF-sponsored metacomputing centers at NCSA and at San Diego. Also, several domain specific repositories are maintained here at the University of Tennessee using the latest version of RIB.

RIB has proven to be a successful application for allowing metadata interoperation across the Internet, which promotes software sharing and reuse. Development of the toolkit has resulted in a

stable release (ver. 2.1) that has undergone testing on IRIX, SunOS 4.x, Windows, AIX, and Linux. While no time frame has been established for the next RIB release, development is currently underway including the addition of many new features, some of which have come at the request of current users. Links to all of the publicly available repositories currently utilizing the RIB toolkit can be found at http://www.nhse.org/RIB/.

The concepts developed in the prototype Digital Library of Mathematical Functions will be applied to the development of a complete Web-based mathematical reference service for special functions to be maintained at NIST. This service will be a model for other highly interactive mathematical reference data projects.

Improvements to the Java language and its environment will be institutionalized via the standardization process. Vendors of Java compilers and JVMs will be required to support such changes, resulting in significant performance enhancements for scientific applications. Standardized numerical APIs will increase the ease of coding, portability, and reliability of Java applications.

List of Publications and Technical Reports

Ronald F. Boisvert, Shirley V. Browne, Jack J. Dongarra, Eric Grosse and Bruce Miller. Interactive and Dynamic Content in Software Repositories. University of Tennessee Computer Science Department Technical Report ut-cs-97-351, February 1997. http://www.cs.utk.edu/~library/TechReports/1997/ut-cs-97-351.ps.Z

R.F. Boisvert and B. Miller, Enhancing the Interactivity of Software and Data Repositories Using Java. in *Proceedings of the IMACS World Congress on Scientific Computing, Modeling and Applied Mathematics*, Berlin, July 1997. ftp://math.nist.gov/pub/reports/boisvert/IMACS97.ps.gz

R.F. Boisvert, J.J. Dongarra, R. Pozo, K.A. Remington and G.W. Stewart. Designing Numerical Libraries in Java. *Concurrency: Practice and Experience*, vol. 10 (No. 11-13), 1998, pp. 1117-1129. ftp://math.nist.gov/pub/reports/boisvert/java98.ps.gz

R.F. Boisvert, J.J. Dongarra, R. Pozo, K.A. Remington and G.W. Stewart. Designing Numerical Libraries in Java. in *Proceedings of the 1998 Workshop on Java for High Performance Computing*, Palo Alto, CA, March 1998.

Shirley Browne, Jack Dongarra, Jeff Horner, Paul McMahan, Scott Wells. Technologies for Repository Interoperation and Access Control. University of Tennessee Computer Science Department Technical Report ut-cs-98-395, April 1998. http://www.cs.utk.edu/~library/TechReports/1998/ut-cs-98-395.ps.Z

Shirley Browne, Paul McMahan and Scott Wells. Repository in a Box Toolkit for Software and Resource Sharing. University of Tennessee Computer Science Department Technical Report ut-cs-99-424, May 1999. http://www.cs.utk.edu/~library/TechReports/1999/ut-cs-99-424.ps.Z

Java Numerics Working Group. Improving Java for Numerical Computation http://math.nist.gov/javanumerics/reports/jgfnwg-01.html, October 1998.

Java Numerics Working Group. Recent Progress of the Java Grande Numerics Working Group, http://math.nist.gov/javanumerics/reports/jgfnwg-02.html, June 1999.

D.W. Lozier, B.R. Miller and B.V. Saunders. Design of a Digital Mathematical Library for Science, Technology and Education. in *Advances in Digital Libraries, Proceedings of the ADL'99 Conference*, Baltimore, May 1999. http://math.nist.gov/DigitalMathLib/publications/nistir6297.ps.gz

J. Millar, P. McMahan, and J. Dongarra. RIBAPI - Repository in a Box Application Programmer's Interface. University of Tennessee Computer Science Department Technical Report ut-cs-00-438, January 2000. http://www.cs.utk.edu/~library/TechReports/2000/ut-cs-00-438.ps.Z

Keith Moore, Shirley Browne, Jason Cox and Jonathan Gettler. Resource Cataloging and Distribution System. University of Tennessee Computer Science Department Technical Report ut-cs-97-346, January 1997. http://www.cs.utk.edu/~library/TechReports/1997/ut-cs-97-346.ps.Z

Bibliography

Digital Library of Mathematical Functions
http://math.nist.gov/DigitalMathLib/

Guide to Available Mathematical Software (GAMS)
http://math.nist.gov/gams/

Harness
http://www.epm.ornl.gov/harness/

Java Numerics
http://math.nist.gov/javanumerics/

Matrix Market
http://math.nist.gov/MatrixMarket/

NetSolve
http://www.cs.utk.edu/netsolve/

Repository In A Box (RIB)
http://www.nhse.org/RIB/

Resource Cataloging and Distribution System (RCDS)
http://icl.cs.utk.edu/projects/rcds/index.html

Scalable Networked Information Processing Environment (SNIPE)
http://www.netlib.org/utk/projects/snipe/

SONAR
http://icl.cs.utk.edu/projects/sonar/index.html